The Impact of Varying Levels of Consistency and Coaching on Deception Detection Accuracy of Children's Injury Reports

By Cassy C. Gillingham (Thesis) submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science in Experimental Psychology (Developmental Psychology)

Department of Psychology, Faculty of Science

Memorial University of Newfoundland

May 2023

St. John's Newfoundland and Labrador

Abstract

The present study examined factors that impact the believability of children's testimonies. There are both forensic and real-world implications for the believability of children's testimonies in court and their given word. The present study aimed to examine transcripts of 8-10 year-olds who were coached or not in order to determine the impact of adding and removing inconsistencies on veracity assessments made by the general population (N= 1,383). It was hypothesized that consistency would account for the increased believability of coached reports. The study found that inconsistency impacted participants' veracity decisions, with a greater number of less consistent transcripts being rated as untruthful. Coaching did not impact the believability of transcripts when consistency was considered. Overall, the hypothesis that consistency accounts for coaching's increase in believability was supported.

Keywords: Consistency, coaching, believability, veracity, children

Acknowledgements

I would first and foremost like that thank my supervisor Carole Peterson for her unwavering support and guidance throughout this process. Without Carole's persevering support on this project the completion of my thesis would not have been possible.

Secondly, I would like to thank all of those supportive individuals who helped make this project a reality. Penny Voutier, lab assistant, for her preparation of data files and meticulous data keeping. As well, many thanks to all the honour's students for their work collecting, entering data and creating stimuli: Kailey Roche, Michelle Tapp, Sarah Williams and Sydney Greene. In addition, all of the MUCEP students who aided in data collection and study promotion.

A special thanks to Darcy Hallett whose guidance in statistical analysis has been essential in bringing this project to fruition, as well as my committee members for their patience and support all of these years.

Finally, I would like to thank my family and faith in God for encouraging me over the course of my research. For their support and encouragement I am forever grateful and owe all my current and future success. "Commit to the Lord whatever you do, and he will establish your plans" (*New International Version*, 2011, Prov. 16:3)

Table of Contents

List of Tables

The Impact of Varying Levels of Consistency and Coaching on Deception Detection

Accuracy of Children's Injury Reports

Children have had a complicated history when testifying in court, and as such, are not always viewed as credible eyewitnesses by a jury (Bala, 2016; Bruck et al., 1998; Orbach & Lamb, 1999). However, whether child eyewitnesses are viewed as credible is crucial in cases where a child's testimony may be the sole piece of evidence for a jury to consider. This often happens in cases of sexual assault where only the perpetrator and victim are present during the crime (Bala, 2016). If a child's credibility is impacted by factors unrelated to their veracity it is conceivable that jurors will make inaccurate judgements about that child's truthfulness in court. In worst-case scenarios, this can lead to inaccurate verdicts of guilt or innocence, where the accused are wrongly imprisoned or the perpetrator walks free.

Although much research has been conducted on factors that increase credibility in adult eyewitnesses, little research exists about the factors that impact children's credibility in court (Warren et al., 2015). Two potential factors are the consistency of the children's reports and coaching. Both are investigated in the present study, along with how they interact.

In the current study, consistency is defined as the extent to which an account provides details that agree with what was said earlier in that same account. Consistency is one of the most commonly cited factors that increases credibility in adults within research asking participants to judge the veracity or truthfulness of eyewitness testimonies (Deeb et al., 2018; Klemfuss & Ceci, 2012). Yet, little research on the topic of consistency has been conducted in regard to child eyewitnesses (Fisher et al., 2013). Studies also show that coaching is a factor that increases children's credibility (Warren et al., 2015). Coaching is when a child is aided by an adult in how to present their side of the story, prior to giving testimony. Research shows that laypersons are

1

more likely to believe children who are coached, regardless of whether they are telling the truth or are fabricating a lie (Warren et al., 2015). The interaction between consistency and coaching is also of key importance, because it is possible that the reason children's coached reports are so believable is that coaching increases consistency in children's testimonies irrespective of whether they are truthful.

The current study examined transcripts of children's accounts of a time they were injured and brought to the emergency room. This was meant to mimic the statement a child might give to police. Half of the children reported a veridical account of their injury experience while the rest were fabricating (i.e., describing a fictitious event that never occurred). In addition, half of each group of children were coached by their parents before their interview. Transcripts of these interviews have been used in other research (Peterson et al., 2011; Warren et al., 2015). In the present study, a third of these transcripts were altered to become more consistent (i.e., some extant inconsistencies removed), a third were altered to become less consistent (i.e., inconsistencies were added), and a third remained unchanged. This unique combination of transcriptions allowed for an examination of the impact of parental coaching and within-report consistency on children's credibility, an area where there has been little research to date.

The following section reviews past research that served as background for the current study, starting with children's history of testifying in court and their credibility or believability in the eyes of a jury. This is followed by a review on lie-detection ability in adults and adults' veracity judgements of children's reports. This includes a discussion about conflicting literature on the topic of adults' experience with children in general, and law enforcement's ability to determine children's veracity. As well, there is a brief review of the impact a juror's confidence has on the accuracy their judgement. Next, factors that lower the credibility of testimonies are

2

compared, specifically the impact of consistency between and within reports. This is followed by a summarization of the impact of parental coaching on children's believability. Finally, there is a brief summary of research surrounding the use of written transcripts and their validity in the context of the current study.

Children's History of Testifying in Court

Canadian court systems prior to the 1980s considered child abuse to be a rare event (Sas et al., 1996). Beginning in the 1970s concerns began to arise that child sexual abuse was much more prevalent than previously believed, as workers began to uncover cases of child sexual abuse in clinical practice (Sas et al., 1996). As a result, in 1981 the Canadian Government formed the Committee on Sexual Offences against Children and Youth, which became known as the "Badgley Committee" (Sas et al., 1996). The findings from this committee mobilized significant changes within the court system when dealing with children and sexual assault.

One major finding of this committee was that child eyewitnesses were not being seen as reliable and were therefore kept from testifying in court (Bala, 1993, as cited in Sas et al., 1996). The younger the child, the less credible they were seen to be and therefore the less likely they were to be allowed to testify in court (Sas et al., 1996). In addition, judges were required to forewarn jurors of children's unreliability as witnesses, undermining their credibility (Bala, 1993, as cited in Sas et al., 1996). These legal "safeguards" made justice in regard to children's testimonies nearly unachievable, and led to child sexual assault being viewed as "rare" due to low conviction rates. After examining the current state of the legal system in regard to children's sexual assaults the Badgley Committee came up with 52 recommendations. In response to these recommendations Bill C-15 was passed in 1988 and Bill C-126 in 1993. These Bills put forward a number of

amendments that paved the way for children to testify in court, resulting in an increased likelihood of justice for victims of child sexual assault (Bala, 1993, as cited in Sas et al., 1996).

In response to this dramatic change in the court system, that allowed children to testify, there was a surge of research about children's credibility while testifying in court during the 1980s. At the time, researchers found two predominant biases held by jurors. The first is that children are generally truthful, since they are naive (Goodman et al., 1984). The second is that children are easily suggestible and therefore unintentionally provide inaccurate testimonies (Goodman et al., 1984).

Even after these changes, competency evaluations continued to center on inquiring if children understood abstract concepts, such as the meaning of an oath, promise and truth. This traditional competency evaluation was confusing and time consuming for children with short attention spans (Lee et al., 2010). Research demonstrated that it often resulted in children being unable to testify, when they were capable of giving their testimony (Lee et al., 2010).

In 2006 the *Canada Evidence Act* was implemented as a result of psychological research. This represented a shift to more evidenced-based competency evaluations of children's testimony, allowing more children to testify in court (Lee et al., 2010). The Act was introduced as a guide for a child under the age of 14 when testifying in court.

The Canada Evidence Act requires a testifying child to promise to tell the truth, but does not ask them to explain the concept of telling the truth or other abstract concepts (Lee et al., 2010). Judges assess whether a child is capable of answering questions about the event in question, which is established based on videos of initial interviews taken by police prior to their court date (Lee et al., 2010). Therefore, if a child is testifying in court, a judge has already made the decision for the jury that a child is able to comprehend and accurately answer basic interview

questions, that this child witnessed the event in question and can recall the event. Once children are in court they are assumed to be competent if they promise to tell the truth. However, if there is an issue with competency it falls on the "party who challenges the child's capacity" to prove the child is unable to understand questions relating to the event in question (Lee et al., 2010). If there are issues with competency then a judge should ask questions pertaining to the child's age, name, school, etc. to demonstrate whether the child can answer questions and allow the jury to witness the child's verbal skills (Lee et al., 2010). From this point on, it is up to a jury to decide whether a child is credible. Credibility is an accuracy judgment about whether a testimony is perceived as accurate and truthful. Research shows that a child's credibility is impacted by many factors. Two important ones, and the focus of the present study, are consistency of testimony and whether a child is coached. Other factors include whether a child appears honest, confidence of the child, and their perceived suggestibility (Bruck et al., 2002; McCauley & Parker, 2001; Ruva & Bryant, 2004; Shao & Ceci, 2011). At the present time, some adult beliefs about children's credibility fall into the category of common myths or assumptions, and therefore adults' judgments about credibility are still not reliable ways of determining if a child is telling the truth or a lie in forensic contexts (Klemfuss & Ceci, 2012).

Deception in Adults

From the beginning of time, humans have been interested in the topic of lie-detection (Vrij, 2008a). Questions that have been explored are, "How accurate are we at detecting the deceptions of others?" and "How can we become more accurate at detecting deception?" Recent meta-analyses estimate that adults' truth and lie-detection abilities are around 54%, essentially no greater than chance (Bond & Depaulo, 2008). Therefore, adults are not proficient at lie-detection in general.

When examining different subsets of the adult population there is evidence of a truth-bias for those who do not have experience with lie-detection as part of their everyday careers (Vrij, 2008b). When law enforcement officers were asked to watch a video of strangers and report their veracity, their accuracy was 55.9% (Vrij, 2008b). In comparison, when laypersons (university students) were asked to make the same judgements of veracity their accuracy rate was 54.3% (Vrij, 2008b). These overall rates of lie-detection are nearly identical; however, differences in these populations emerged when examining for truth bias. Laypersons showed a truth bias (63.4% accuracy for truthful accounts and 48.2% accuracy for untruthful accounts) meaning they were more accurate in rating truthful accounts and less accurate in rating deceptive accounts. In contrast, law-enforcement officers did not show a truth-bias (56.1% accuracy for deceptive accounts and 56.4% accuracy for truthful accounts). It appears that laypersons are more likely to label truthful statements as true and duped into believing false statements, whereas law-enforcement officers are equally suspicious of truth-tellers and lie-tellers (Vrij, 2008b).

In conclusion, the overall accuracy rates for adults demonstrate that there is no significant difference in lie-detection abilities found in these so called "lie-detection" experts as a whole (just above 50% accuracy rates) (Bond & Depaulo, 2008).

Adults' Lie-Detection of Children

Although many meta-analyses find that adults are chance level lie detectors when determining the veracity of other adults' testimonies, this is not necessarily true of adults judging children's testimonies (Gongola et al., 2017). However, a 2017 meta-analysis found that indeed, adults are only chance level lie-detectors when examining children's reports (Gongola et al., 2017). Parallel to findings in studies about adults, Quas et al. (2007) found that a truth bias can also be seen when adults rated children's true accounts.

As for children's lie detection by professionals who work with children, there are divergent findings in the literature. With increasing experience working in a field related to children, some studies demonstrate a gradient, where those who worked more closely with children for longer periods of time were better than those with less experience at detecting children's lies (Crossman & Lewis, 2006). Crossman and Lewis (2006) adapted a modified temptation resistance paradigm that depicts the concealment of a transgression. This particular kind of deception is designed to keep the child or someone the child is familiar with out of trouble, similar to what is seen in forensic situations. Preschoolers were asked to sit facing a mirror, while the experimenter brought a toy into the room. The toy was placed out of view and children were instructed "not to peek" as the experimenter left the room. Afterwards, children were asked if they peeked, thus giving them the opportunity to conceal their own transgression. In this study, judges were shown videos of this paradigm and were asked if children were being truthful or untruthful in their peeking answers. In general, these judges were below chance levels at detecting children's veracity, adding to a host of other studies stating that adults are poor lie detectors when it comes to children (Bond & DePaulo, 2006; Quas et al., 2017). In contrast, those professionals who worked with children could detect lies at chance levels, showing an improved lie-detection ability compared to that seen from the general public (Crossman & Lewis, 2006). However, their ability still seemed to be no greater than chance accuracy in general; this finding has been supported in other reports (e.g., Reinhard et al., 2011). When examining professionals who have experience working with children (e.g., teachers), evidence shows a truth bias that is greater than that of the general public. One explanation for this seems to be the nature of the teacher-student relationship where trust is important (Reinhard et al., 2010). In comparison, in other professions where a lie-bias or "investigator bias" can be seen

(e.g., police officers), trust in a suspected criminal would be counter-intuitive (Meissner & Kassin, 2002; O'Sullivan et al., 2009). However, there are also other studies that found no difference in lie detection ability between police officers', social workers', teachers' and the general public's ability to successfully detect children's lies (Amodt & Custer, 2006; Vrij et al., 2006).

Confidence in one's judgments is also a factor that has been explored. The evidence converges on the conclusion that individuals who are confident in their judgements are no better at lie detection than those with little confidence (Leach et al., 2014). In other words, confidence has no significant correlation with accuracy (Leach et al., 2004). Indeed, certain professions who are more confident about their judgments are actually less accurate when detecting children's lies, for example customs officers (Leach et al., 2004).

With so many divergent findings in the literature, the explanation could be that different studies vary in their experimental procedures. Many studies examining lie-detection use a brief video and ask participants to judge veracity (O'Sullivan et al., 2009). This has little similarity to what professionals are faced with in the real world. For example, police officers generally have access to multiple interviews and other evidence before making a decision about an individual's truthfulness. A teacher knows their individual students after spending hours of time with them, better than a complete stranger (as is often shown in experiments - Reinhard et al., 2009). Furthermore, studies with higher accuracy rates for professionals have been demonstrated when the conditions of the research reflect both the emotional and cognitive nature found within professionals' occupations (O'Sullivan et al., 2009). Therefore this finding is present in conditions where face validity (the study appears to measure what it is said to measure) and differential validity (accurately measuring the differences between groups) is high (O'Sullivan et

al., 2009). Finally, meta-analyses have few studies about professionals from which to generalize lie-detection ability (Amodt & Custer, 2006). The most common professionals being studied are police officers, but studies examining students' ability to detect veracity far outnumber those about professionals. When examining lie-detection ability, it is possible that scenarios with high face and content validity for a student have very little face and content validity for a professional. This could be one possible explanation for the diverse findings within the literature.

Since studies about adults' abilities to detect children's lies vary so greatly, we must look to meta-analyses in order to get a wholistic picture. A meta-analysis by Gongola et al. (2017) examined 45 studies about adults evaluating children's veracity. This study found adults to be slightly better than chance (54.34%) at determining veracity, with elevated levels of accuracy for true statements (a truth bias) whereas false statement accuracy remained at chance levels. These findings are nearly identical to what is found in the adult literature (Gongola et al., 2017). However, when breaking the analysis into age subcategories, adults were better at detecting veracity in younger children. For children between the ages of 3-5 adults made correct judgements 60% of the time, for ages 6-9 they were accurate 56% of the time and for older children, ages 10-15 they were accurate 52% of the time. Additionally, professionals outperformed laypersons at veracity judgements; however this was a very small percentage (56% vs. 54%) and not a big enough difference to have real-life or forensic implications. Finally, the literature on adults detecting children's lies shows a small but significant effect for confidence. This suggests that those who are more confident may be slightly better at children-lie detection. However, this effect size is so small it is unclear if there is any practical significance.

In sum, although adults are slightly above chance levels at detecting lies told by children, these levels do not differ from their ability to detect adults' lies (i.e., 54%). Research suggests professionals may have a slight advantage, as well as those who are more confident in their assessments. However, these trends exist in the adult literature as well and do not differentiate the adult from child lie-detection literature. Furthermore, the younger the child, the greater the likelihood of accuracy by adult assessors in determining veracity. These age-related rates seem to be one of the few differences in the literature, with children ages 3-5 being significantly less likely to deceive adults with their falsehoods. Finally, there is a truth-bias found in the children's literature, suggesting that adults perceive children as less likely to lie in general.

Impact of Consistency

Consistency is considered to be high when multiple pieces of information agree with each other and are not contradictory. This consistency can be seen in a forensic context through between and within-statement consistency (Vredevelt et al., 2014). Between statement consistency is when different statements, taken from interviews at two separate times, align in meaning and facts (Vredeveldt et al., 2014). In contrast, within-statement consistency is when facts align within a statement made during one interview (Vredeveldt et al., 2014). The current study examines within-statement consistency. Consistency in general is seen as one of the main ways jurors decide if a witness is being truthful, in particular child witnesses (Deeb et al., 2018; Klemfuss & Ceci, 2012). Consistency, or agreement within statements, is seen by both professionals and laypersons as an indication of truthfulness (Desmarais et al., 2009). In comparison, inconsistencies within a statement are often interpreted to mean that a witness is concealing information or being untruthful (Vredevelt et al., 2014).

Consistency can be seen as a quick and valid way to determine the veracity of a suspect or witness; this is known as the consistency heuristic (Deeb et al., 2018; Vrij et al., 2010). However, research shows that looking at consistency alone may not always lead to an accurate validity assessment (Brubacher et al., 2019; Fivush et al., 2002; Quas et al., 2007; Vredeveldt et al., 2014). Upon first thought, it makes sense that liars would be less consistent than truth tellers. However, studies show that oftentimes liars are actually more consistent than those providing an honest statement and this phenomenon has been demonstrated in studies with both adults and children (Brubacher et al., 2019; Granhag & Stromwall, 2002; Stromwall & Granhag, 2005; Stromwall and Jonsson, 2003; Quas et al., 2007). One theory for why consistency can be inaccurate when used as a heuristic is because of the repeat versus reconstruct hypothesis, proposed by Granhag and Stromwall (1999). It is theorized through this hypothesis that liars understand that inconsistent statements will reveal their deception and therefore they practice and prepare their statements before being interviewed. This results in a consistent testimony, with very little in the way of perceived corrections, omissions, or contradictions (all forms of inconsistencies). However, when a witness is truthful about their statement they recall said statement from memory. Unfortunately, one thing we know about memory from research is that it is imperfect. Thus, truthtellers may omit things they simply missed or forgot and these facts can be remembered later (Fischer et al., 2009). Indeed research with adult witnesses demonstrates that we are more likely to be inconsistent if there is a delay between the experienced event and the interview, if we poorly encoded the experience, if there is a delay between a subject's first interview and follow-up interviews or if they are trying to correct aninaccuracy that emerged from a leading question or coercion (Vredevelt et al., 2014).

Importantly, research on how children report their experiences indicate that there are even more reasons why a child may give an inconsistent account of an event (Brubacher et al., 2019). One instance is when children are interviewed multiple times about one event. Studies show that children are likely to add new details over multiple interviews (Fischer et al., 2009; La Rooy et al., 2009). This is because children often answer open ended questions with multiple, partial answers and do not give long and complete answers (Poole & Dickinson, 2011). This means a child may add new information over the course of multiple interviews resulting in low between-statement consistency. However, this does not mean a child is being untruthful, it is simply an artifact of how a child responds to open-ended questioning. Part of the reason why this can lead to inconsistencies is that interviewers will cut off questioning prematurely and this leads to that child telling new details in a later interview (Brubacher et al., 2019).

Another reason why children can become inconsistent is if a child becomes bored with questioning (Brubacher et al., 2019). People of all ages can go off topic when they become uninterested. However, because of the understanding gap between a child and adult, an off-topic conversation can come across to the interviewer as a child being inconsistent about their story. This can also be seen when children are involved in multiple scenarios that are similar in nature, but not the same situation. For example, if a child has broken a bone more than once, that child may amalgamate two separate, yet similar instances, into one story (Price et al., 2016). Indeed interviewers should be vigilant about clarifying that the child has not switched topics so as not to misunderstand details about an event.

Another way that children may seem inconsistent in their reports is through detail errors (Brubacher et al., 2019). Reports will often contain errors about particular details and when these errors do not impact the overall coherence of a report it does not necessarily mean a child is

being untruthful (Brubacher et al., 2019). The younger the child, the more likely the child is to slip up and give these inaccurate details. These inconsistencies are often innocuous errors that arise when children are trying to remember a specific event from multiple similar events or because a child thinks this is something that would typically happen and incorporates it into their story. Examples of this include "source errors" where a child may recall contextual information from multiple sources and combine it into one narrative (Foley, 2014). For these reasons, experts recommend not paying close attention to contextual details that are unessential to the story (Brubacher et al., 2019).

Finally, focused questions can sometimes lead children to give contradictory answers from what they originally stated in an open-ended question (Brubacher et al., 2015). When this happens, research advises that the context of the narrative is considered before disregarding a testimony based on an apparent contradiction (Brubacher et al., 2019). Reasons for this are that when a child is asked a focused question, their understanding of that question may limit their answer. Two examples where this happens are: when an interviewer uses unnecessary words, and if an interviewer uses a word/words the child may be unfamiliar with (Brubacher et al., 2019).

Despite consistency not always being an accurate indicator of lie-detection, some literature suggests that there are instances where consistency can be a helpful tool for determining truthfulness (Bruck et al., 2002). This applies for interviewing methods that increase cognitive load (Bruck et al., 2002). Cognitive load is derived from the cognitive load hypothesis, where a liar experiences an increased cognitive load when coming up with details about their statement because they are deriving it internally and not directly from an experience. In contrast, a truthful subject experiences less cognitive load because they simply have to recall the events

that took place (a relatively simple task). For example, one way interviewers increase cognitive load is by asking the individual being interviewed to tell their story backwards or start from the end and work their way back to the beginning. This added difficulty increases "cognitive load", or the number of variables a liar has to keep track of. In this case, both fabricating a story and simultaneously trying to recall their fabricated story backwards. This would increase the task difficulty for a liar, whereas someone telling the truth would be able to easily recall their story from memory (Bruck et al., 2002). However, techniques that increase cognitive load are unethical to use with young children due to the fact that they may be too complicated for both truthful and untruthful children (Vrij et al., 2017). Hence, they are not addressed further in the current study.

Another important question is how inconsistencies are viewed, by both members of the general public as well as professionals such as police officers, prosecutors, and judges. It is important to note that although consistency is seen as key by some people in determining whether a suspect is truthful or not (Deeb et al., 2018; Vrij et al., 2010), these beliefs are not often studied empirically (Deeb et al., 2018). However, some studies show that laypersons believe inconsistencies demonstrate deceit in any and all circumstances, whereas professionals hold a more complicated view (Cashmore & Trimboli 2006; Deeb et al., 2018).

Laypersons are quite skeptical of inconsistencies in victims and eyewitnesses (Malloy & Lamb, 2010). In particular, when children testify, jurors are increasingly skeptical of their testimony when there are inconsistencies. These inconsistencies may arise due to a number of reasons, such as delays between the event and initial interview, incomplete accounts, general inconsistencies or if an account was retracted (Quas et al., 2005). A study by Cashmore and Trimboli (2006) examined jurors' perceptions of child eyewitnesses and concluded that when a

juror perceived a child's testimony as consistent, it was correlated with their belief that the child was a highly credible witness.

Another place where perceptions of consistency are impactful is within the courtroom. Here, views are influenced by prosecution and defense lawyers. Their views on consistency can be seen through their questioning of child witnesses. A study by Denne et al. (2020) found that consistency was essential to establish when interviewing victims of alleged child sexual assault as consistency was routinely addressed in questions by prosecutors (79% of total questions asked of the child witness) and defense attorneys (89% of total questions asked of the child witness). Lawyers address a child's consistency within the courtroom through questions and in turn these questions can help establish a child's consistency in the mind of the jury. This is particularly impactful when we consider that judges are more likely to mention inconsistency in their rulings on acquittals than in convictions (Connolly et al., 2009)

Although research demonstrates that consistency and credibility are highly correlated in the minds of the jury and even professionals, little is known about how variation of consistency impacts assessments of a child's credibility (Cashmore & Trimboli, 2006; Deeb et al., 2018). This includes whether consistency operates on an all-or none-principle, or if there is a threshold for the amount of inconsistency individuals can handle before the story becomes unbelievable. To my knowledge, there are no studies that address the importance of varying levels of consistency in children's testimony. The current study manipulates varying amounts of consistency in otherwise identical transcripts allowing the changing importance of consistency to be examined at high and low levels.

Impact of Coaching

Studies show that coaching can increase children's believability for both true and false reports (Gongola et al., 2017; Warren, 2015). In fact, even giving a child time to prepare lies by themselves can make their reports more believable (Stromwall, 2007). However, the impact of coaching has not been demonstrated with children of all ages. Younger children, particularly children between the ages of 3-5, may be more difficult to successfully coach (Vrij et al., 2002; Warren, 2015). In one study by Vrij et al. (2002), the researchers were able to get strangers to successfully coach children on CBCA (Criteria-Based Content Analysis) criteria when children were above the age of 7. They were not able to successfully coach children younger than this. Originally, the researchers attempted to coach children ages 5-6; however, the children became very confused by their instructions and the heavy coaching condition had to be cancelled for this age group. In agreement with this finding, Warren et al. (2015) found that lies told by children ages 5-7 were detected at chance levels whether they were coached or not by parents. However, 8-14 year-olds were more likely to be believed when they were coached in both truthful and fabricated conditions. Thus, it appears that coaching becomes increasingly effective with older children. For these older children, coaching is a factor that consistently increases their believability (Warren et al., 2015). Coaching then can be seen as a positive way to help increase the believability of children's interviews (Warren et al., 2015). This is useful when children are telling the truth. However, when children are being deceitful, coaching is a negative factor that can decrease a layperson's accuracy of veracity judgements (Talwar et al., 2006). Indeed, the current body of literature seems to conclude that children can be coached to tell false reports, and these false reports are normally longer, more detailed and more believable than uncoached reports (Talwar et al., 2006; Warren, 2015).

So why does coaching make a child's testimony so believable, in particular among children older than 8 years of age? To date there are few studies on the topic of coaching, and fewer still on why coaching increases believability. However, there seems to be the belief that children, like adults, find lying to be cognitively taxing (Talwar et al., 2017). Preparing and coaching children before they give a testimony reduces the cognitive load and makes it easier to give their statement. Additionally, practicing beforehand also decreases inconsistencies in children's testimony, which may contribute to why coached reports are more believable (Warren et al., 2015). A study by Warren et al. (2018) examined whether linguistic analysis could shed light on why coached children are more believable. Using linguistic inquiry and word count software (LIWC), transcripts of injured children (a stressful event) were analyzed for differences in linguistic trends for coached versus uncoached children. The study found that coached children are more likely to have longer reports, use more insight terms, negations and show more negative emotions when telling a distressing story (Warren et al., 2018). Thus it is possible that coached reports are more believable because they are longer and more detailed than uncoached reports.

In light of these theories about the underlying reasons for the effectiveness of coaching, a fundamental issue to consider is whether inconsistencies should be used to determine if a testimony is accurate (Klemfuss & Ceci, 2012). However, some studies found consistency was not predictive of statement accuracy (Brubacher et al., 2019; Hammond & Fivush, 1991). In a study by Quas et al. (2007) where children were coached to tell a lie about being touched (in a non-harmful manner), it was found that those children who told the truth ended up being the least consistent (in comparison to those who successfully produced a coached lie). This study

demonstrates that when children are coached to lie, those rehearsing a lie, especially with the help of an adult, are actually more consistent than those telling the truth.

In conclusion, it would be inaccurate to assume that simply because a witness provides a consistent testimony, that they are being truthful. As well, it is erroneous to assume that just because a witness is inconsistent that they are a liar. Instead, witnesses may be both inconsistent and truthful, especially if they are very young children (Brubacher et al., 2019; Hammond & Fivush, 1991). Furthermore, the presence of inconsistencies may be influenced by coaching, where a liar can be consistent within and between their statements (Quas et al., 2007). Therefore, using consistency as a test for veracity is problematic. When considering the consistency of a testimony, individuals should focus on the situation surrounding the testimony given, in conjunction with consistency, to get the most accurate veracity assessment (Deeb et al., 2018; Quas et al., 2007).

Transcripts in Forensic Context

In preparation for court cases, interviews and testimonies are often transcribed and these transcripts are used throughout the judicial process (Warren et al., 2015). For example, transcripts may be used during jury deliberation, as well as being reviewed by police, defense, prosecution, and expert witnesses. They are useful in that they allow comparison between accounts. Therefore, they are often at the center of research as both a method of testimony presentation and an area of study (Warren, 2015; Wojciechowski et al., 2017).

 Studies demonstrate that when individuals pay attention to non-verbal cues (ex: body language/ what the individual does) they are actually less accurate than those who pay attention to verbal cues (ex: what the witness says or what is written in a transcript) when determining

veracity (Vrij, 2008). Those who examine only non-verbal cues are more likely to hold a "lie bias," meaning that the judge is more likely to rate a video of an individual testifying as a "liar" if they ignore verbal content (ex: details said or written in a transcript) altogether (Bond & DePaulo, 2006). This results in higher rates of false accusations (Vrij, 2008). Reasons for a lie bias come from common assumptions about liars versus truth tellers. These assumptions are commonly held beliefs that have little basis in reality. One belief assumes that liars are nervous when being interviewed, whereas truthtellers are not (Vrij & Granhag, 2007).

In one analysis that compared mediums of testimony, researchers examined individuals' ability to distinguish truthful accounts vs. deceptive ones (Bond & DePaulo, 2006). This analysis compared videos with no audio, videos with audio, only audio, and transcripts. When comparing these mediums, videos with no audio led to significantly higher rates of inaccurate veracity judgements in comparison to the other mediums, which did not differ from each other (Bond & DePaulo, 2006). This study showed impairments in lie-detection when only visual cues are taken into consideration.

This leads us to question the cues to which judges pay the most attention. Should they focus on visual or non-visual cues when both are present? Also, can or should judges consider both equally? Mann et al. (2004) found that judges who focus on inconsistencies and amount of detail, or verbal content cues, tend to be better lie-detectors; in contrast, those who are poorer at lie-detection tend to focus more on visual tells, such as gaze aversion. This has similarly been seen in other studies where those asked to make veracity judgements were less accurate when citing non-verbal cues (Porter et al., 2007; Vrij, 2008; Vrij & Mann, 2001). Overall, those who look for non-verbal cues to truthfulness are often duped. To add to this effect, stereotypes about "liars" encourage untrained and trained individuals to focus on nonverbal cues (Vrij et al.,

19

2019). This leads to a general misconception about what a liar "looks" like. This misconception is that a liar has a specific look or visual tell that marks them as distinct from those telling the truth (Vrij et al., 2019). In reality, research shows that liars often appear no different than those telling the truth (Vrij et al., 2019). Furthermore, these tells vary based on the situation and individual (Vrij et al., 2019).

If a focus on verbal cues leads to better lie-detection, is there any impairment when non-verbal and verbal cues are presented at the same time, such as in a video? Research suggests that in cases where individuals are not experts in the area of judgement, when there are not multiple statements to compare, and when there is a lack of extraordinary statements, verbal cues are essentially ignored for non-verbal cues (Granhag & Stromwell, 1999; Mann 2004). Those who focus solely on these non-verbal cues are at a disadvantage when detecting lies, in comparison to those who focus on verbal cues. Therefore, in the current study we used only transcripts, so as to remove the error associated with non-verbal cues (Hartwig et al., 2011; Vrij, 2008).

The current study sought to analyze consistency of testimony. Here, consistency referred to details within a singular account that did not contradict each other. For example, a child stating their mother picked them up after school and then later stating their father picked them up after school would be an inconsistent account. Transcripts were used because verbal indicators were of interest; since transcripts contain only verbal indicators, judges had to focus solely on these cues. This design allowed us to examine judges' tendency to use verbal cues in detecting deception without the interference of non-verbal cues (Hartwig et al., 2011; Vrij, 2008b). The design of the current study manipulated consistency. Some transcripts were made more consistent, others made less consistent, and yet others were kept unchanged. Videos would not allow for this type of manipulation because children would have to be videotaped and coached to

tell more consistent versus less consistent stories. This may be possible when brief scenarios are described by children but would be nearly impossible when extensive interviews are involved, as was the case here. Therefore, the ease with which transcripts could be manipulated, the fact that they are used in forensic contexts, and that they allow raters to use only verbal clues, make transcripts an appropriate choice for the present study.

The Present Study

Upon examining the literature to date a few key points were evident. The first was that adults are not skilled at lie detection (Vrij, 2008b). Secondly, adults are equally unskilled at detecting children's lies, unless they are told by very young children (Quas et al., 2017; Warren, 2015). In fact, adults' credibility judgements are no more accurate than chance. Finally, there are many poorly understood factors that impact perceptions of children's credibility; two of which being coaching and consistency. However, these factors have not been studied extensively. Do lay judges, for example, know whether individuals use the consistency heuristic when making accuracy judgements? Furthermore, can consistency explain why children's coached stories seem so believable? The current study's unique set of transcripts allowed these questions to be explored.

In the current study, children's true and false accounts (half of the children were coached and half uncoached) were made more or less consistent or they were left unchanged. University students and members of the public were then asked to rate the credibility of the children's accounts.

The study was guided by the following hypotheses:

1. Participants will be chance level lie-detectors.

2. There will be a truth bias in the data.

3. Coached testimonies will be more likely to be believed than uncoached testimonies.

4. Removing inconsistencies will result in judgements of higher credibility, independent of the actual truthfulness of children's accounts.

5. Allowing significant inconsistencies will result in judgments of lower credibility, independent of the actual truthfulness of the children's accounts.

6. The effect of coaching on credibility will disappear when consistency is added to the model.

7. There will be no relationship between accuracy and assessment when judges have experience working with children.

8. Confident judges will be no more accurate than chance in their judgements of children's accounts.

Method

Each participant was given a transcript of a child describing the time they were sent to the emergency room for a serious injury (e.g., breaking a bone) or a transcript of a child fabricating an account of injury and hospital treatment.

There were three between-subject variables: 1. Whether the child was actually telling the truth or a lie, (i.e., the transcript's veracity - Truth or Lie); 2. Whether the child was coached by an adult (Coached or Uncoached); 3. Whether the transcript was manipulated to be less or more consistent or left unaltered (Less Consistent or More Consistent or Original Transcript). Additionally, the transcripts were evenly divided by the nested variable of child's gender (Male or Female). Gender was not included in the analysis since it has not been found to be significant in previous studies; however half of all transcripts were male and half were female to avoid it being a confounding variable (Warren et al., 2015). This created a 2 (veracity: truth or lie) x 2 (coaching: coached or uncoached) x 3 (manipulation: less consistent or more consistent or original) between-subjects factorial design.

Three versions of each transcript were used or created in order to explore the factor of consistency: (a) the original transcript, (b) a more consistent account (three inconsistencies present in the children's original accounts were eliminated if possible), and (c) a less consistent account (three additional inconsistencies were added). Inconsistency examples included child assertions of mom taking them to the hospital early in the interview, but later stated that dad took them to the hospital, or inconsistent identity of friends at different places in the interview, or inconsistent descriptions of where their injury took place, etc. These are the sorts of

inconsistencies that currently exist in the transcripts, and in the manipulation, the number of such inconsistencies were increased, decreased, or remained the same.

Participants

There were 1,421 people who began the study. However, the participant pool that was used offered the option provided of "research observation" (i.e., participants may take part in the study without having their results submitted); therefore this combined with some individuals failing to complete the survey led to 38 responses being discarded. Participants were given the option of two incentives, to be entered into a draw for a $100 gift card or to be given 1 credit towards a psychology course they were currently taking (applicable only to students). The final sample (n = 1383) was 76.1% female, 22.5% male and 1.4% who did not select any option for their gender, i.e. missing information (n = 19). The average age of participants was 25.45 years (SD = 10.09). The ethnicity of the sample was mainly Caucasian (67.2%). The largest percentage of the sample was current university students (45.3%). However, members of the general public were included (23.5%), as were participants who did not indicate whether or not they were currently a student (31.2%).

Materials

Original Transcripts

This study took advantage of an extant corpus of transcripts that were collected in the course of other research (Warren et al., 2015). The interviews were previously audio recorded and transcribed verbatim, with the exception of omitting any identifying information.

These interviews were then typed into transcripts to create four separate groups of children for Warren et al.'s (2015) study. For children in the truthful-uncoached condition,

researchers used audio recordings from previously gathered child-interviews (Peterson, 2010). The interviewed children were treated at the emergency room of a local children's hospital for an injury. Triage nurses directed interviewers to appropriate patients and parental consent was given for an interview. Truthful-coached reports were gathered from children who had experienced an injury similar to those from the emergency room, but had been coached by their parents before being interviewed by a researcher. The third group of children represented the fabricated-uncoached group, where children were recruited from students' connections (i.e., friends and family) as well as from a child-care centre. These children were asked to create an untrue story about an injury similar to those of children who were recruited from the local emergency room. This group was unassisted when creating their story. The final group was a fabricated-coached condition, where children were also asked to create an untrue story about an injury, again similar to those children who actually experienced the injury, but these children were coached by their parents for four days leading up to the interview.

Appropriate consent was gained from the parents of all children to take part in these interviews and for the transcripts to be used in future research. All aspects of the prior study were approved by the Interdisciplinary Committee on Ethics in Human Research of Memorial University. For the present study, appropriate verbal consent was obtained from Dr. Kelly Warren and Dr. Carole Peterson for use of the transcripts.

In each interview, children were asked to provide free recall details of their injury and hospital treatment. Next they were asked a set of open-ended questions, followed by direct questions to clarify statements. Ages and genders of the children were listed at the top of the transcripts for participants to see.

Manipulation of Existing Transcripts

To ensure the validity of this manipulation, the researchers examined the original transcripts found in Warren et al.'s (2015) study to determine the types of inconsistencies that typically occurred. The original responses from that study were used to see if participants picked up on any inconsistencies. The different types of inconsistencies participants stated influenced their decision making were noted and manipulated in the current study. Therefore, the inconsistencies added here are ecologically valid and naturally occurring inconsistencies.

Manipulating the current transcripts involved a team of researchers. Originally, we took 32 transcripts from children aged 8-10 years of age. Children aged 8-10 were chosen because coaching was shown to be most effective in increasing credibility in this age group in the study by Warren et al. (2015) where the transcripts were initially used. The 32 transcripts (half male and half female) were then made to be more consistent, made less consistent, or were kept the same (no manipulation) creating a total of 96 different transcripts: 32 being more consistent, 32 less consistent, and 32 original transcripts. This was done by removing up to three or as many as were possible to remove (More Consistent Conditions) and adding three inconsistencies (Less Consistent Conditions). A breakdown of the number of inconsistencies removed to create the subset of more consistent transcripts can be found in Table 1 as these conditions varied in the amount taken away. It is important to note that some transcripts were already so consistent that it was impossible to remove three inconsistencies. As a result, each condition within the more consistent transcripts (Coached-True, Coached False, Uncoached True and Uncoached False) had varying amounts of inconsistencies within them. For example, one Coached-True transcript may have had no inconsistencies, where another may still have had one inconsistency. Therefore, the commonality within conditions was that an effort was made to add, remove or leave

consistencies unaltered. In the "More Consistent" manipulation the transcripts with no inconsistencies were left in their original state and therefore were essentially the same as "original transcripts" since they were already as consistent as possible. This was particularly an issue in the "Coached-False" condition (see Table 1).

Table 1

Amount of Inconsistencies Removed in More Consistent Transcript Conditions

More Consistent Transcript Conditions	Inconsistencies Removed			
	3 removed	2 removed	1 removed	0 removed
Coached-True	3	3	2	0
Coached-False	1	0	1	6
Uncoached-True	2	2	1	3
Uncoached-False	2	4	1	1

The team we used to add the inconsistencies (less consistent conditions) and take them away (More consistent conditions) included research assistants and graduate students as well as a professor who went through the original 32 transcripts and identified inconsistencies. Then as a team decisions were made of where to add or take inconsistencies away.

Experimental Procedures

Recruitment

This study was approved by the university's Interdisciplinary Committee on Ethics in Human Research. To recruit participants, research assistants went into undergraduate psychology

courses and discussed the need for participants in a new study, the PREP website where students can go and browse studies that allow them to get extra credit towards their psychology courses was used, posters and pamphlets were left around the university explaining the need for participants, and links were posted on social media sites, such as Facebook. Recruitment materials explained that the rationale of the study was to look at factors that make a child witness credible; however, the true hypotheses (regarding consistency and coaching) of the study were not fully explained to avoid influencing participants' responses.

Procedure

Participants who agreed to take part in this study were randomly assigned to 1 of the 96 transcripts using either an automated program (i.e., when on either Survey Monkey or Qualtrics) or drawing from a random number table (i.e., when in person). There were two waves of recruitment. Participants who took part in the initial wave of recruitment from 2015-2017 used the Survey Monkey website or were given a hard copy survey package. The Survey Monkey website randomly assigned online participants to conditions. However, the random assignment meant that there was an uneven number of participants in each condition, with some having far fewer participants than others. In order to fix this problem a second wave of online recruitment took place in 2018 using the Qualtrics website. Using this website, participants were able to be randomly assigned to transcripts that had a lower number of participants. Once the transcript had the desired number of participants they were removed from the pool of randomly assigned transcripts. This was unable to be done using Survey Monkey and therefore was replaced by Qualtrics in the second wave of online recruitment.

Those who took part in 2018 were surveyed using the Qualtrics website[1]. Students who took part in this study were surveyed in a computer lab at Memorial University spaced far enough apart from other participants that they could not see others' computer screen or papers. Those recruited from the general public completed their survey online wherever they were using their electronic devices. Consent was obtained at the outset of the study. Participants were then asked to read a transcript of an interview describing the time a child was injured and had to go to the emergency room in order to decide if the child was telling the truth or a lie. After the participants read the transcript they were asked to complete a survey package assessing the veracity of the child's statement. Dr. Carole Peterson and Dr. Kelly Warren developed the original survey and it was revised for this study. Specifically, participants were asked whether the child was telling the truth (Yes or No), how confident they were in their answer on a likert scale from 1-5 (1 being *not at all confident* and 5 being *very confident*), and why they came to this conclusion. The reasoning behind the participants' decisions also acted as a manipulation check. This way we could see if participants explicitly reported consistency as a reason for the children's credibility. Additionally, the survey consisted of a demographics section, a section of questions about experience the participant may have with children in general, children as eyewitnesses, and traumatic injuries. In this section participants were asked if they had any relevant experience with children or forensic situations. Jobs considered to involve experience with children were teachers, babysitters, parents or daycare workers. Participants were also asked if they had any jobs related to forensic experience; these jobs included social workers, lawyers, and police officers.

---

[1] Switching from Survey Monkey to Qualtrics allowed participants to be randomly assigned to conditions, while simultaneously allowing researchers to fill each condition with a certain number of participants.

Once participants completed the survey, they were given a debriefing form to explain the true intent of the study and to tell them how their data would be used. In-lab participants were given a hard copy, as well as an online copy, which they could take with them from the lab. In contrast, online participants were given a form which they could print or even take a picture of on their own device. This form explained to participants that some amount of deception was needed since in a jury situation, individuals never truly know if the eyewitness is giving a truthful account. Additionally, the true hypotheses were revealed and it was explained that researchers did not want to influence participants' decisions by providing the hypotheses earlier in the study.

Results

General Overview of Accuracy

Frequency statistics were completed to determine the overall rate of accuracy for all combined conditions (49.82% accurate and 50.18% inaccurate). In order to compare participants' ability to accurately judge transcripts to chance (score of .50) a one sample t test was conducted. Participants were no more accurate ($M = 0.50$, $SD = 0.50$) than chance, $t(1382) = 0.13$, $p = .893$, $d = 0.004$, 95% CI [-.02, .03].

Next, the possibility of a truth-bias was assessed for the overall sample. In order to compare the likelihood of participants' truth or lie decisions (a score of 1.5) a one sample t test was conducted. Participants were no more likely ($M = 1.48$, $SD = 0.50$) to rate the transcripts as truthful, $t(1382) = -1.59$, $p = .113$, $d = -0.04$, 95% CI [-.05, .01].

Frequency statistics were used to determine accuracy rates for each condition; these are outlined in Table 2, followed by one sample t tests to compare each condition's accuracy to chance.

Table 2

Accuracy Percentages for Participants in Each Condition

Manipulation	Condition			
	Coached-True	Coached-False	Uncoached-True	Uncoached-False
Original Transcripts (OG)	60.36%	34.23%	55.49%	48.95%
Less Consistent Transcripts (LC)	36.47%	64.95%	29.03%	64.55%
More Consistent Transcripts (MC)	62.77%	36.27%	63.22%	42.39%

Original Transcripts' Accuracy

In order to compare participants' ability to accurately judge transcripts to chance (a score of .50) a one sample *t* test was conducted on each condition for the original transcripts. Participants in the coached-true condition were significantly more accurate ($M = 0.60$, $SD = 0.49$) than chance, $t(110) = -2.22$, $p = .028$, $d = -0.21$, 95% *CI* [-.20, -.01]. Those in the coached-false condition were significantly less accurate ($M = 0.34$, $SD = 0.48$) than chance, $t(148) = 4.04$, $p < 0.001$, $d = 0.33$, *CI* [.08, .24]. Participants in the uncoached-true condition were not significantly different ($M = 0.55$, $SD = 0.50$) than chance, $t(172) = -1.45$, $p = .149$, $d = -0.11$, *CI* [-.13, .02]. Finally, those in the uncoached false condition were not significantly different ($M = 0.49$, $SD = 0.50$) from chance, $t(189) = 0.29$, $p = .773$, $d = 0.02$, *CI* [-.06, .08].

Less Consistent Transcripts' Accuracy

Secondly, one sample *t* tests were conducted on each condition for the less consistent transcripts. Participants in the coached-true condition were significantly less accurate (*M* = 0.37, *SD* = 0.48) than chance (a score of 0.50), *t*(84) = 2.58, *p* = .012, *d* = 0.28, 95% *CI* [.03, .24]. Those in the coached-false condition were significantly more accurate (*M* = 0.65, *SD* = 0.48) than chance, *t*(96) = -3.07, *p* = .003, *d* = -0.31, *CI* [-.25, -.05]. Participants in the uncoached-true condition were significantly less (*M* = 0.29, *SD* = 0.46) than chance, *t*(92) = 4.43, *p* < .001, *d* = 0.46, *CI* [.12, .30]. Finally, those in the uncoached false condition were significantly more accurate (*M* = 0.65, *SD* = 0.48) from chance, *t*(109) = -3.17, *p* = .002, *d* = -0.30, *CI* [-.24, -.05].

More Consistent Transcripts' Accuracy

Thirdly, one sample *t* tests were conducted on each condition for the more consistent transcripts. Participants in the coached-true condition were significantly more accurate (*M* = 0.63, *SD* = 0.49) than chance (a score of .50), *t*(93) = -2.55, *p* = .013, *d* = -0.26, 95% *CI* [-.23, -.03]. Those in the coached-false condition were significantly less accurate (*M* = 0.36, *SD* = 0.48) than chance, *t*(101) = 2.87, *p* = .005, *d* = 0.28, *CI* [.04, .23]. Participants in the uncoached-true condition were significantly more accurate (*M* = 0.63, *SD* = 0.49) than chance, *t*(86) = -2.54, *p* = .013, *d* = -0.27, *CI* [-.24, -.03]. Finally, those in the uncoached false condition not significantly different (*M* = 0.42, *SD* = 0.50) from chance, *t*(91) = 1.47, *p* = .145, *d* = 0.15, *CI* [-.03, .18].

Factors That Impact Judgement

Initially, a binomial logistic regression was conducted with participants' judgement of the transcripts as the outcome variable and actual veracity of the transcript (true or false) and whether the child was coached (coached or uncoached) as the predictor variables. The present

study was based on Warren et al's 2015 study and the enter method was selected based on the results of the 2015 study. All variables were entered into the model concurrently. Variables entered into the model included the actual veracity of the transcript and whether or not the child was coached. The goodness-of-fit test, the Hosmer-Lemeshow, was not significant ($p > .05$) which indicated that the model was correctly specified. Other statistical tests of importance were the *-2 Log likelihood = 1909.39* and *Nagelkerke R squared* = 0.01. The binomial regression analysis found the variable of actual veracity was non-significant ($p > .05$). However, the variable of coaching was found to be significant. Therefore, when actual veracity and coaching were entered into the model, coaching was found to contribute to the model, B = -0.25, *SE* = 0.11, *Wald* = 5.32, p = 0.021. The estimated odds ratio showed a decreased likelihood of participants who judged the uncoached children judging them as truthful, *Exp(B)* = 0.78, 95% *CI* [.63, .96]. Thus, the child's actual veracity had little effect, while coaching increased the likelihood of judging a child as truthful.

A second binomial logistic regression was completed with participants' judgement of the transcripts as the outcome variable and manipulation of the transcript (original, less consistent or more consistent transcripts), actual veracity (true or false) and whether the child was coached (coached or uncoached) as predictor variables. The enter method was used to enter the following variables: manipulation of transcripts, actual veracity of transcripts, whether the child was coached, as well as all 2-way and 3-way interactions. The variables were selected based on the results of Warren et al's 2015 study, with our new variable (manipulation and interactions) being added to test our present hypotheses. All predictor variables were entered concurrently. The Hosmer-Lemeshow goodness-of-fit was found to be non-significant ($p > .05$) which means that the model was correctly specified. Other statistics of interest were the *-2 Log likelihood =*

1833.59 and *Nagelkerke R squared* = 0.08. The analysis revealed the variables of actual veracity, whether the child was coached, 2-way and 3-way interaction terms were non-significant ($p >$ .05). However, the variable of manipulation was found to contribute to the model, B = -0.95, *SE* = 0.30 , *Wald* =10.33 , p = .001. The estimated odds ratio showed a decrease in the likelihood of participants judging the children with less consistent transcripts as telling the truth in comparison to those children with original transcripts, *Exp(B)* = 0.39 , 95% *CI* [.22, .69]. The results are outlined in table Table 3.

Table 3

The Effect of Manipulation, Coaching and Actual Veracity on Predicting Judgement

Variable	B	SE	Wald	p	Exp(B)	95% CI for Low	95% CI for High
Coached	-0.18	0.25	0.52	.471	0.84	.52	1.36
Veracity	0.24	0.26	0.88	.349	1.28	.77	2.13
Manipulation (ref: OT)			14.47	.001			
LC	-0.95	0.30	10.33	.001	0.39	.22	.69
MC	0.12	0.29	0.19	.666	1.13	.64	1.99
Manipulation (ref: MC)							
OT	-0.21	0.20	1.16	.281	0.81	.55	1.19
LC	-1.25	0.22	31.84	.000	0.29	.19	.44

Note: OT= Original Transcript, MC= More Consistent, LC= Less Consistent, ref= reference category

When less consistent (LC) transcripts (coded as 1) and original transcripts (OT) (coded as the reference category) were compared, a negative coefficient suggested that the less consistent manipulation was significantly more likely to result in a child being assessed as telling a lie than a child was in the original transcripts (B = -0.95, SE = 0.30, p < 0.001).

The second dummy variable compared the more consistent (MC) manipulation transcripts (coded as 2) and the original transcripts (OT) (the reference category), and showed a positive coefficient; however, the difference was not significant ($B = 0.12$, $SE = 0.29$, $p = .666$). The transcript's actual veracity (true or false) was a non-significant predictor for assessment ($B = -0.003$, $SE = 0.11$, $p = 0.98$).

A final comparison of the less consistent (LC) manipulation (coded as 2) and the more consistent manipulation transcripts (MC) (coded as the reference category) showed a negative coefficient that suggests the less consistent manipulation was significantly more likely to result in a child being assessed as telling a lie than a child in the more consistent transcripts ($B = -1.25$, $SE = 0.22$, $p < 0.001$).

Overall, when looking at judgements of children's stories the child's veracity was not shown to be a significant factor in predicting assessment. Coaching was originally shown to be a significant factor, but once consistency was added to the model, the variance could be accounted for by the consistency manipulation. Therefore, consistency was a significant factor in predicting assessment and explained the variance attributed to coaching in the first binomial regression. Post-hoc tests determined that less consistent transcripts were significantly more likely to be assessed as lies compared to original transcripts and compared to more consistent transcripts. Finally, more consistent transcripts were not significantly more likely to be assessed as true compared to original transcripts.

In order to examine the impact of not being able to take away consistencies for some transcripts in the "More Consistent" condition, a third binomial logistic regression was completed as a manipulation check. All transcripts that were unable to have any consistencies

removed for the consistency manipulation were removed from the data. This resulted in the pool

of transcripts shrinking from 96 to 66. Overall, a total of thirty transcripts and their respective

data were removed. In the analysis participants' judgement was the outcome variable and

manipulation of transcripts, actual veracity of transcripts, coaching, 2-way and 3-way

interactions were the predictor variables. Again, the Hosmer-Lemeshow goodness-of-fit was not

significant ($p > .05$) indicating the model was correctly specified. Other statistics included the *-2

Log likelihood* = 1292.29 and *Nagelkerke R squared* = 0.07. The results were similar to the

previous regression analysis. The predictor variables of actual veracity, coaching, 2-way and 3-

way interaction terms were not significant ($p > .05$). Also similar to the original findings, the

variable of transcript manipulation was found to contribute to the model, B = -0.98, *SE* = 0.30 ,

Wald = 10.76 , p = .001. The estimated odds ratio showed a decrease in the likelihood of

participants judging the children with less consistent transcripts as telling the truth in comparison

to those children with original transcripts, *Exp(B)* = 0.38 , 95% *CI* [.21, .68]. These results are

presented in table Table 4.

Table 4

The Effect of Manipulation, Coaching and Actual Veracity on Predicting Judgement with the

Removal of Thirty Consistent Transcripts

Variable	B	SE	Wald	p	Exp(B)	95% CI for Low	95% CI for High
Coached	-0.18	0.27	0.44	.509	0.84	.49	1.42
Veracity	-0.17	0.33	0.24	.622	0.85	.44	1.63
Manipulation (ref: OT)			14.74	< .001			
LC	-0.98	0.30	10.76	= .001	0.37	.21	.68
MC	0.10	0.29	0.12	.724	1.11	.63	1.95
Manipulation (ref: MC)							
OT	-0.10	0.29	0.12	.72	0.90	.51	1.59
LC	-1.08	0.31	12.05	< .001	0.34	.19	.63

Note: OT= Original Transcript, MC= More Consistent, LC= Less Consistent, ref= reference

category

The first comparison of the less consistent (LC) manipulation transcripts (coded as 1) and

the original transcripts (OT) (coded as the reference category) showed a negative coefficient that

suggests the less consistent manipulation was significantly more likely to result in a child being

assessed as telling a lie than a child was in the original transcripts (B = -0.98, SE = 0.30, p = .001).

The second dummy variable compared the more consistent (MC) manipulation transcripts (coded as 2) and the original transcripts (OT) (the reference category), and showed a positive coefficient; however, the difference was not significant (B = 0.10, SE = 0.29, p = .724).

A final comparison of the less consistent (LC) manipulation (coded as 2) and the more consistent manipulation transcripts (MC) (coded as the reference category) showed a negative coefficient that suggests the less consistent manipulation was significantly more likely to result in a child being assessed as telling a lie than a child was in the more consistent transcripts (B = -1.08, SE = 0.31, p < .001).

Overall, removing transcripts that were completely consistent did not change the significance of any results and the manipulation was shown to be effective.

Expertise with Children and the Law

Pearson correlations were completed to assess the possible relationship between degree of experience with children and accuracy. There was no significant relationship found between amount of experience and the accuracy ratings of teachers (r = 0.08, p = 0.084, n = 466), daycare workers (r = 0.01, p = 0.879, n = 500), babysitters, (r = 0.003, p = 0.944, n = 700) or parents (r = -0.006, p = 0.897, n = 525).

Similarly, Pearson correlations were completed to assess the possible relationship between degree of forensic experience and accuracy. There was no significant relationship found

between level of experience and accuracy for social workers (r = -.03, *p* = 0.528, n = 407), lawyers (*r* = -0.09, *p* = 0.066, *n* = 390), or police officers (*r* = -0.06, *p* = 0.220, *n* = 447).

Confidence and Accuracy

The frequencies of participants' confidence ratings are reported in Table 5 (*M* = 3.60, *SD* = 0.80). A Pearson correlation of participants' confidence in their assessment of children's reports showed there was no significant correlation between confidence and accuracy (*r* = 0.01, *p* = 0.704, *n* = 1379). Finally, a one sample *t* test was ran to compare confident participant's (those who rated confidence a 4 or 5) accuracy to chance (a score of .50). Confident participants were found to be no more accurate (*M* = 0.50, *SD* = 0.50) than chance, *t*(768) = 0.14, *p* = .885, *d* = 0.01, 95% *CI* [-.07, .08].

Manipulation Check

The question about "why" participants made their judgements was noted by researchers. Consistency was frequently mentioned (the majority of participants referred to consistency when explaining their judgements) and therefore the manipulation of consistency in testimony was deemed effective.

Table 5

Frequencies of Confidence Ratings

Confidence Rating	Frequency
1 - Not at all Confident	10
2	82
3 - Somewhat Confident	519
4 -	611
5 - Very Confident	157

Discussion

Overall, this study found that adults are chance level lie detectors when it comes to assessing the true and false accounts of children around the ages of 8-10. Furthermore, when these children are coached by a parent, they are significantly more likely to be believed than when they are uncoached. This holds true for when children are being truthful and for when they are being deceitful. As hypothesized in this study, the increase in children's believability when they are coached can be explained by within-testimony consistencies. In contrast, within-testimony inconsistencies were associated with a decrease in children's believability.

This study was completed as a follow-up of Warren et al.'s (2015) study that examined coaching in children ages 5-14 who discussed a true or fictional event about breaking a bone or getting lacerated that resulted in a trip to their local children's hospital Emergency Room. One of the findings of that study was that when children of ages 8 to 10 years were coached by parents on how to tell their story, there was a significant increase in the believability of the children's accounts by laypersons. Furthermore, this increase was seen in cases where they were both telling the truth and telling a lie. Researchers hypothesized that one reason these coached children were so believable was that they were more consistent than those children who were not coached by their parents. The current study set out to examine this hypothesis by manipulating coached and uncoached children's testimony to be more consistent or less consistent than their original testimony. This study found the hypothesis that highly consistent testimonies account for an increase in these children's believability, even when coaching is included in a regression model.

Overall Accuracy

43

Participants' overall accuracy when determining children's veracity for the present study was at chance levels (49.80% accurate). Although these rates were slightly below those in meta-analyses examining adult's ability to judge children's reports (54%) they nevertheless add to the literature that suggests adults are chance level lie detectors when it comes to children (particularly between the ages of 8-10). A meta-analysis by Quas et al. (2017) found that the younger the child, the more accurate adults were at detecting their lies, with lies by younger children (around the age of 3) being detected with a rate of 60% accuracy. This suggests that the reason for the small discrepancy in our findings compared to those of Quas et al. (2017) is that Quas et al. had a wider age range that included children as young as 3 years of age as well as children up to age 10 years whereas the present study examined only relatively older children (8-10 years). For these older children, Quas et al. (2017) also found adults' accuracy to be closer to chance or 50% accuracy.

In Warren et al.'s (2015) study it was found that children who were coached and telling the truth were most accurately judged by adults. This was similarly seen in the current study when comparing the transcripts from the original study. When transcripts were manipulated by increasing consistency, judgment accuracy of these true accounts became even higher and this was not impacted by the variable of coaching. Therefore, highly consistent transcripts were seen as more believable. Additionally, when transcripts were altered to be less consistent, accuracy increased for the deceitful conditions independent of the coaching variable. This led to inconsistent transcripts being judged as less believable. This shows what studies call the "consistency heuristic", that is the tendency for individuals to assume that because a testimony is consistent it is true and to be more likely to believe that an inconsistent testimony is false (Deeb et al., 2018; Vrij et al., 2010).

44

Impact of Consistency on Believability

Contrary to what was hypothesized, the gradient of consistency did not impact children's believability in the manipulated transcripts. Transcripts manipulated to be more consistent were not significantly more likely to be believed, even after removing ten transcripts in the "More Consistent" condition. In contrast, transcripts manipulated to be less consistent led participants to be significantly less likely to believe the child. This finding fits with past research that suggests laypersons are skeptical of inconsistencies and label consistent testimonies as truthful (Cashmore & Trimboli, 2006; Quas et al., 2005). The present study focuses on single-report consistencies or consistencies within a single transcript. This is unique in comparison to previous studies since these studies focus on between-report consistency (Cashmore & Trimboli, 2006; Deeb et al., 2018). In fact, to my knowledge there are no studies to date that examine the effect of consistency within children's reports. Therefore, there are important conclusions that can be drawn from the present study that cannot be made from previous studies. One is that within-report inconsistencies have the power to influence a layperson's decision about veracity. Indeed, inconsistencies are more important to predicting a layperson's veracity decision than whether the report is actually true or false. Furthermore, the effect of consistency does not seem to be additive, in that as long as the report was not glaringly inconsistent there was no difference in believability. Inconsistent reports were significantly less likely to be believed.

Impact of Coaching on Believability

The current study was motivated by a prior study by Warren et al. (2015) that had examined the impact of parental coaching on a cohort of children between 5 and 14 years of age. In Warren et al.'s study, coaching was shown to be most effective with the 8 to 10 year age

group. In this cohort, parental coaching increased children's believability regardless of whether the accounts were true or fabricated. Consequently, that age group was chosen for the present study so that a more detailed examination of the effect of coaching could be done.

Other studies have also shown that coaching increases believability in children's reports (Lyon et al., 2008; Vrij et al., 2004). The current study suggests that it is not the act of coaching per se but rather the decrease of major inconsistencies in children's accounts which occurs after coaching that may account for coaching's effect on people's veracity judgments. In the present study, when coaching was assessed after manipulating consistency, it was no longer predictive of veracity decisions made by lay judges. However, what remained predictive was the inconsistent transcripts. Overall, this speaks to the weight that jurors give inconsistency when judging testimony. When coached transcripts were made inconsistent, the impact of coaching became non-significant.

Experience, Confidence and Accuracy

Past literature assessing the relationship between experience and accuracy has shown mixed results. Some studies have suggested that those who have experience with children (such as teachers) have increased accuracy when it comes to labeling their stories as truthful in comparison to those with less experience with children (Crossman et al., 2006). In contrast, others say that experience has no impact (Leach et al., 2004). Furthermore, some studies state that those with forensic experience are better at lie-detection (such as police officers) (Mann et al., 2004). Yet other research suggests that this experience does not lead to an enhanced truth-lie detection ability (Leach et al., 2004). In agreement with the latter research, the present study found no correlation between jobs that included experience working with children (i.e., teachers,

daycare workers, parents or babysitters) and accuracy of veracity ratings. As well, there was no correlation between forensic experience (i.e., the occupational experience of police officers, lawyers and social workers) and accuracy of veracity decisions. Thus, the present research agreed with those that found no relation between experience and accuracy. Overall, this shows that depending solely on the judgements of "experts" to determine the testimonial accuracy of children (in the absence of corroborating evidence) would be a mistake.

The present study also agrees with past literature that found confidence has no significant correlation with accuracy (Leach et al., 2004). Overall, although participants were confident in their decisions with an average confidence rating of 3.6 out of 5, with 5 being *very confident*, participants were no different from chance in their accuracy decisions. This finding demonstrates that although individuals may be confident in their accuracy decisions, this belief has little to no impact on whether they actually made a correct veracity assessment. This is important to note when considering real-life implications since a juror may feel confident in their veracity decision, but there is no evidence to support this being associated with an increased ability to make an accuracy judgement.

Forensic Implications

In the present study, coaching was found to be a significant factor related to the believability of children's testimony. These findings are worrisome because of the fact that coaching increases the likelihood a child will be believed independent of whether they are telling the truth or not. Therefore, in a forensic situation a child could be believed or discredited on factors that are unrelated to truthfulness. Indeed, if the goal is to increase correct veracity assessments by jurors, it is possible that coaching could be an issue. Estimates for intentionally

fabricated accusations of child maltreatment are rare (4% of all cases); however when cases are related to custody or access, this number jumps to 12% (Trocmé & Bala, 2005). False accounts of maltreatment reported by children are also rare, representing only 2% of all false reports (Trocmé & Bala, 2005). Despite these cases being infrequent, they still represent a portion of cases making their way through the justice system. Faller (2007) examined professionals' beliefs about coaching. In this study, approximately 80% of professionals believed that they had encountered at least one case of coaching throughout their career. Although it is unknown how frequently children are coached, it is clear it is encountered by professionals and represents a portion of intentionally false reports in the court system. Overall, the present study shines a light on an important topic that must be addressed in the 21st century.

When examining children's testimony, there are many reasons why a child may be inconsistent besides the testimony being inaccurate (Brubacher et al., 2019). Despite this, much importance is placed on the consistency of testimonies in forensic situations (Deeb et al., 2018; Klemfuss & Ceci, 2012). Andrews et al. (2015) examined transcripts of children (up to the age of 12) who were questioned, crossed examined and even re-cross examined as alleged victims of child abuse. Transcripts were coded to identify any "self-contradictions" (i.e., where children's testimonies contradicted something they had previously said). It was found that in 95% of the testimonies, children contradicted themselves and this was consistent in all age categories (6-9 year-olds and 10-12 year-olds). Furthermore, these contradictions were elicited by the questioning of both prosecutors and defense attorneys. Although this study found at least one contradiction in 95% of transcripts, there was an average of approximately 5 inconsistencies in direct examinations and an average of approximately 12 inconsistencies in cross examination.

This aligns with previous studies that also found that the majority of children questioned in court produced inconsistencies (Zajac & Canon, 2009; Zajac et al., 2003).

There is a growing concern about whether justice can be served when methods used to elicit the truth seem to elicit contradictions that allow children's credibility to be unfairly drawn into question (Andrews et al., 2015; Denne et al., 2020). Denne and colleagues (2020) examined the questions of prosecutors and defense attorneys in Childhood Sexual Abuse (CSA) cases and found that 79% of prosecutors asked questions about consistency, as did 89% of defense attorneys. Both sides addressed the child's consistency through questioning more than any other area of questioning when trying to establish or undermine credibility. Furthermore, defense attorneys were more likely to ask developmentally inappropriate questions undermining children's credibility than were prosecutors (Denne et al., 2020). Studies show that children's reports in general demonstrate a number of inconsistencies (Brubacher et al., 2019). In addition to this, children's reports can be particularly inconsistent when children have experienced multiple similar events (Price et al., 2016). Researchers examining inconsistencies in children's testimonies have found that children who experienced multiple events were often more inconsistent than children who experienced a single event (Price et al., 2016). Furthermore, 88% of children who experienced multiple events reported at least one inconsistent response across two interviews. In these cases, inconsistencies were most often a result of children reporting the similar event that they had experienced instead of the "target event" identified in questioning. Therefore, it is highly likely that a child who experienced multiple episodes of abuse could appear inconsistent to a jury, but still be telling the truth (Price et al., 2016). In conclusion, children's reports are fraught with inconsistencies and these inconsistencies tend to increase when various factors that are part of the court system come into play, such as suggestive

questioning, close ended questions, source monitoring errors, and multiple experienced events (Andrews et al., 2005; Brubacher et al., 2019; Faller, 2005; Price et al., 2016). The overall result is highly inconsistent testimonies within a forensic context.

Moving forward, the present study allows those in the general population, researchers and those who study law to take note of their preconceived notions related to consistency when making veracity decisions in the case of children. This is important since consistency does not equate with truth-telling and vice-versa, especially when it comes to children who may have added reasons for having an inconsistent story (Brucbacher et al., 2019). Implications for everyday use for adults who interact with children are that it is useful to know that an inconsistent story does not make a child a liar (Brubacher et al., 2019), whether this be in the principal's office, at home, or at the babysitter's house. Additionally, when it comes to previous research on coaching, one reason why coaching is so effective with children ages 8-10 is that it decreases the inconsistency of their testimony. Another question that this study alludes to is, "Is the reason very young children's testimonies are not improved by coaching that they are still inconsistent, even after parental coaching?" However, perhaps most importantly, the present study illustrates exactly why it is useful for judges to educate jurors on the types of errors children make when giving testimonies and why these mistakes do not mean they are being deceitful, as shown in past research (Brubacher et al., 2019; Fivush et al., 2002; Quas et al., 2007; Vredeveldt et al., 2014). The current study allows us to see that jurors do take children's inconsistency to heart and this factor is correlated with the final decisions they make as to whether or not they believe a child. This finding has dire implications when extrapolated to the sensitive cases where children find themselves testifying in a North American court system

where children's history of being believed by adults has been anything but straightforward (Bala, 1993 as cited in Sas et al., 1996; Sas et al., 1996).

Limitations

One limitation in this study is that only within-report consistency was studied. Research shows that although within-report consistency is important for jurors and professionals in their decision-making processes, there is also between-report consistency to be considered (Deeb et al., 2018). Between-report consistency examines how two separate testimonies agree with each other and are not contradictory. There has been research to suggest that laypersons use between-report consistency when making veracity decisions (Granhag & Stömwall, 2001). Furthermore, it is often the case that children provide multiple accounts about a single event (Brubacher et al., 2019). This creates the opportunity for between-report inconsistencies. Although the impact of these inconsistencies is not examined in the present study, this would be an important next step for research that assesses factors that impact the believability of child testimonies.

A second limitation is the age of the children providing testimony for the stimuli. In the present study, children in transcripts ranged from ages 8-10 years old. This research is a follow-up study of Warren et al.'s (2015) where the transcripts of 8-10 year-old's believability was shown to be significantly impacted by parental coaching. Interestingly, the current study found that consistency of children's testimony accounted for this increase in believability. However, due to the limited age categories examined, it is difficult to generalize these findings to other age groups, specifically younger children. Does a consistent testimony also result in increased believability for younger children? Warren et al.'s (2015) study seems to suggest that coaching did not increase younger children's consistency and this is possibly why coaching these groups

was ineffective. However, it is impossible to draw these conclusions as those age groups were not studied. Another follow-up study will be needed to address the aforementioned question.

A final limitation of the current study is that a portion of participants took part in the study using an online link using their own personal devices. This could be on their mobile device, personal laptop or other devices. Those who took part in the study through their personal devices may have had outside influences and distractions. Although this allowed us to increase our sample profile and access an increased number of participants, ranging in different ages and populations, it also allows the opportunity for outside influences to impact the research. The impact of this method of data collection is unknown at this time. However, with online programs being increasingly popular for data collection, it is no longer an uncommon issue in the world of research. Finally, the present study is an experimental study and therefore the transferability of results should be interpreted cautiously. Experimental studies have been used in the past to reform court procedures in Canadian Law regarding children's testimony (for example those studies that led to *The Canada Evidence Act* in 2006 (Lee et al., 2010), and particularly those researchers involved in *The Child Witness Project* (Lee et al., 2006). Future research should aim to examine the impact of increased inconsistency in real-world forensic contexts in order to demonstrate transferability of these concepts.

Conclusions

In conclusion, variation in the amount of inconsistency in children's accounts did impact lay judges' ratings of the account credibility. Specifically, greater inconsistency in children's accounts made lay jurors significantly more likely to judge children's accounts as false. In contrast, more consistent transcripts were more likely to be rated as true, showing that amount of

consistency or inconsistency impacts believability. However, these findings were not additive, in that completely consistent transcripts were not rated as more truthful than mildly consistent transcripts. Furthermore, these findings were independent of whether the children were actually truthful or deceitful. The presence of coaching, although originally contributing to lay judges' perceptions, became non-significant when consistency was considered. Therefore, inconsistency actually accounted for coaching's impact on lay judges' veracity decisions. This suggests that the reason coaching is so impactful on veracity judgements is because it decreases inconsistency in children's testimony. Other factors such as experience with the law, with children or participant confidence did not impact these findings.

References

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? *Forensic Examiner*, *15*(1), 6-11.
https://www.proquest.com/scholarly-journals/who-can-best-catch-
liar/docview/207648142/se-2?accountid=12378

Andrews, S. J., Lamb, M. E., & Lyon, T. D. (2015). Question types, responsiveness and self-
contradictions when prosecutors and defense attorneys question alleged victims of child
sexual abuse. *Applied Cognitive Psychology*, *29*(2), 253-261.
https://doi.org/10.1002/acp.3103

Bala, N. C., & Hebert, P. (2016). Views, perspectives and experiences of children in family
cases. https://dx.doi.org/10.2139/ssrn.2887623

Bond Jr, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy
and bias. *Psychological Bulletin*, *134*(4), 477-492. https://doi.org/10.1037/0033-
2909.134.4.477

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and
Social Psychology Review*, *10*(3), 214-234.
https://doi.org/10.1207/s15327957pspr1003_2

Brubacher, S. P., Poole, D. A., & Dickinson, J. J. (2015). The use of ground rules in investigative
interviews with children: A synthesis and call for research. *Developmental Review*, *36*,
15-33. https://doi.org/10.1016/j.dr.2015.01.001

Brubacher, S. P., Peterson, C., La Rooy, D., Dickinson, J. J., & Poole, D. A. (2019). How

    children talk about events: Implications for eliciting and analyzing eyewitness

    reports. *Developmental Review*, *51*, 70-89. https://doi.org/10.1016/j.dr.2018.12.003

Bruck, M., Ceci, S. J., & Hembrooke, H. (1998). Reliability and credibility of young children's

    reports: From research to policy and practice. *American Psychologist*, *53*(2), 136-151.

    https://doi.org/10.1037/0003-066X.53.2.136

Cashmore, J., & Trimboli, L. (2006). Child sexual assault trials: A survey of juror

    perceptions. *BOCSAR NSW Crime and Justice Bulletins*, 20.

    https://www.bocsar.nsw.gov.au/Publications/CJB/cjb102.pdf

Connolly, D. A., Price, H. L., & Gordon, H. M. (2009). Judging the credibility of historic child

    sexual abuse complaints: How judges describe their decisions. *Psychology, Public Policy,*

    and Law, 15(2), 102-123. https://doi.org/10.1037/a0015339

Crossman, A. M., & Lewis, M. (2006). Adults' ability to detect children's lying. *Behavioral*

    Sciences & the Law, *24*(5), 703-715. https://doi.org/10.1002/bsl.731

Deeb, H., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Strömwall, L. A. (2018). Police

    officers' perceptions of statement inconsistency. *Criminal Justice and Behavior*, *45*(5),

    644-665. https://doi.org/10.1177/0093854818758808

Denne, E., Sullivan, C., Ernest, K., & Stolzenberg, S. N. (2020). Assessing children's credibility

    in courtroom investigations of alleged child sexual abuse: Suggestibility, plausibility, and

    consistency. *Child Maltreatment*, *25*(2), 224-232.

    https//doi.org/10.1177/1077559519872825

Faller, K. C. (2007). Coaching children about sexual abuse: A pilot study of professionals'

    perceptions. *Child Abuse & Neglect*, *31*(9), 947-959.

    https://doi.org/10.1016/j.chiabu.2007.05.004

Fisher, R. P., Vrij, A., & Leins, D. A. (2013). Does testimonial inconsistency indicate memory

    inaccuracy and deception? Beliefs, empirical research, and theory. In B. S. Cooper, D.

    Griesel & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness*

    memory, and credibility assessment (pp. 173-189). Springer. https://doi.org/10.1007/978-

    1-4614-5547-9_7

Fivush, R., Peterson, C., & Schwarzmueller, A. (2002). Questions and answers: The credibility

    of child witnesses in the context of specific questioning techniques. In M. L. Eisen, J. A.

    Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp.

    331-354). Taylor & Francis Group. https://ebookcentral-proquest-com.qe2a-

    proxy.mun.ca/lib/mun/detail.action?docID=234086

Foley, M. A. (2014). Children's memory for source. In P. J. Bauer, & R. Fivush (Eds.), *The*

    Wiley handbook on the development of children's memory (pp. 427-452). John Wiley &

    Sons Inc. https://doi.org/10.1002/9781118597705.ch19

Godoy-Cervera, V., & Higueras, L. (2005). Criteria-based content analysis (CBCA) in statement

    credibility assessment. *Papeles del Psicologo*, *26*, 92-98.

    https://www.papelesdelpsicologo.es/English/1249.pdf

Gongola, J., Scurich, N., & Quas, J. A. (2017). Detecting deception in children: A meta-

    analysis. *Law and human behavior*, *41*(1), 44. https://doi.org/10.1037/lhb0000211

Goodman, G. S., Golding, J. M., & Haith, M. M. (1984). Jurors' reactions to child

    witnesses. *Journal of Social Issues*, *40*(2), 139-156. https://doi.org/10.1111/j.1540-

    4560.1984.tb01098.x

Granhag, P. A., & Strömwall, L. A. (2001). Deception detection: Interrogators' and observers'

    decoding of consecutive statements. *The Journal of Psychology*, *135*(6), 603-620.

    https://doi.org/10.1080/00223980109603723

Hamond, N. R., & Fivush, R. (1991). Memories of Mickey Mouse: Young children recount their

    trip to Disneyworld. *Cognitive Development*, *6*(4), 433-448.

    https://doi.org/10.1016/0885-2014(91)90048-I

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the

    detection of deception? A meta-analysis. *Communication Research*, *43*(3), 283-343.

    https://doi.org/10.1177/0093650214534974

Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (2017). Can credibility criteria be

    assessed reliably? A meta-analysis of criteria-based content analysis. *Psychological

    Assessment*, *29*(6), 819-834. https://doi.org/10.1037/pas0000426

Klemfuss, J. Z., & Ceci, S. J. (2012). Legal and psychological perspectives on children's

    competence to testify in court. *Developmental Review*, *32*(3), 268-286.

    https://doi.org/10.1016/j.dr.2012.06.005

La Rooy, D., Lamb, M. E., & Pipe, M. E. (2009). Repeated interviewing: A critical evaluation of

    the risks and potential benefits. In K. Kuehnle & M. Connell (Eds.), *The evaluation of*

child sexual abuse allegations: A comprehensive guide to assessment and testimony (pp. 327-361). John Wiley and Sons Inc. https://psycnet.apa.org/record/2009-01883-014

Leach, A. M., Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). "Intuitive" lie detection of children's deception by law enforcement officials and university students. *Law and Human Behavior*, *28*(6), 661-685. https://doi.org/10.1007/s10979-004-0793-0

Lee, K., Talwar, V., & Bala, N. (2010). The competency of children to testify: Psychological research informing Canadian law reform. *The International Journal of Children's Rights*, 18(1), 53-77. https://doi.org/10.1163/157181809X458544

Malloy, L. C., & Lamb, M. E. (2010). Biases in judging victims and suspects whose statements are inconsistent. *Law and Human Behavior*, *34*(1), 46-48. https://doi.org/10.1007/s10979-009-9211-y

McCauley, M. R., & Parker, J. F. (2001). When will a child be believed? The impact of the victim's age and juror's gender on children's credibility and verdict in a sexual-abuse case. *Child Abuse & Neglect*, *25*(4), 523-539. https://doi.org/10.1016/S0145-2134(01)00224-1

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and human behavior*, *26*(5), 469-480. https://doi.org/10.1023/A:1020278620751

Nahari, G. (2017). Top-down processes in interpersonal reality monitoring assessments. *Psychology, Public Policy, and Law*, *23*(2), 232-242. https://doi.org/10.1037/law0000110

New International Version. (2011). BibleGateway.com. http://www.biblegateway.com/versions/

Orbach, Y., & Lamb, M. E. (1999). Assessing the accuracy of a child's account of sexual abuse:
A case study. *Child Abuse & Neglect*, *23*(1), 91-98. https://doi.org/10.1016/S0145-
2134(98)00114-8

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy:
The effect of lie scenario. *Law and Human Behavior*, *33*(6), 530-538.
https://doi.org/10.1007/s10979-008-9166-4

Peterson, C., Moores, L., & White, G. (2001). Recounting the same events again and again:
Children's consistency across multiple interviews. *Applied Cognitive Psychology*, *15*(4),
353-371. https://doi.org/10.1002/acp.708

Peterson, C., Warren, K. L., & Short, M. M. (2011). Infantile amnesia across the years: A 2-year
follow-up of children's earliest memories. *Child Development*, *82*(4), 1092-1105.
https://doi.org/10.1111/j.1467-8624.2011.01597.x

Poole, D. A., & Lindsay, D. S. (2001). Children's eyewitness reports after exposure to
misinformation from parents. *Journal of Experimental Psychology: Applied*, *7*(1), 27-50.
https://doi.org/10.1037//1076-898X.7.1.27

Poole, D. A., & Dickinson, J. J. (2011). Evidence supporting restrictions on uses of body
diagrams in forensic interviews. *Child Abuse & Neglect*, *35*(9), 659-669.
https://doi.org/10.1016/j.chiabu.2011.05.004

Porter, S., McCabe, S., Woodworth, M., & Peace, K. A. (2007). Genius is 1% inspiration and
99% perspiration… or is it? An investigation of the impact of motivation and feedback on

deception detection. *Legal and Criminological Psychology*, *12*(2), 297-309.

https://doi.org/10.1348/135532506x143958

Price, H. L., Connolly, D. A., & Gordon, H. M. (2016). Children who experienced a repeated

event only appear less accurate in a second interview than those who experienced a

unique event. *Law and Human Behavior*, *40*(4), 362-373.

https://doi.org/10.1037/lhb0000194

Quas, J. A., Thompson, W. C., Alison, K., & Clarke-Stewart, K. A. (2005). Do jurors "know"

what isn't so about child witnesses? *Law and Human Behavior*, *29*(4), 425-456.

https://doi.org/10.1007/s10979-005-5523-8

Quas, J. A., Davis, E. L., Goodman, G. S., & Myers, J. E. B. (2007). Repeated questions,

deception, and children's true and false reports of body touch. *Child Maltreatment*, *12*(1),

60-67. https://doi.org/10.1177/1077559506296141

Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007).

Developmental differences in the effects of repeated interviews and interviewer bias on

young children's event memory and false reports. *Developmental Psychology*, *43*(4), 823-

837. https://doi.org/10.1037/0012-1649.43.4.823

Reinhard, M. A., Dickhäuser, O., Marksteiner, T., & Sporer, S. L. (2011). The case of Pinocchio:

Teachers' ability to detect deception. *Social Psychology of Education*, *14*(3), 299-318.

https://doi.org/10.1007/s11218-010-9148-5

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on

    perceptions of witness credibility and trial outcome. *Journal of Applied Social*

    Psychology, *34*(9), 1919-1944. https://doi.org/10.1111/j.1559-1816.2004.tb02593.x

Sas, L. D., Wolfe, D. A., & Gowdey, K. (1996). Children and the courts in Canada. *Criminal*

    Justice and Behavior, *23*(2), 338-357. https://doi.org/10.1177/0093854896023002006

Shao, Y., & Ceci, S. J. (2011). Adult credibility assessments of misinformed, deceptive and

    truthful children. *Applied Cognitive Psychology*, *25*(1), 135-145.

    https://doi.org/10.1002/acp.1652

Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's

    statements: The impact of a repeated experience on CBCA and RM ratings. *Applied*

    Cognitive Psychology, *18*(6), 653-668. https://doi.org/10.1002/acp.1021

Stromwall, L. A., & Granhag, P. A. (2005). Children's repeated lies and truths: Effects on adults'

    judgments and reality monitoring scores. *Psychiatry, Psychology and Law*, *12*(2), 345-

    356. https://doi.org/10.1375/pplt.12.2.345

Strömwall, L. A., & Granhag, P. A. (2007). Detecting deceit in pairs of children. *Journal of*

    Applied Social Psychology, *37*(6), 1285-1304. https://doi.org/10.1111/j.1559-

    1816.2007.00213.x

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2006). Adults' judgments of children's

    coached reports. *Law and Human Behavior*, *30*(5), 561-570.

    https://doi.org/10.1007/s10979-006-9038-8

Talwar, V., Crossman, A., & Wyman, J. (2017). The role of executive functioning and theory of mind in children's lies for another and for themselves. *Early Childhood Research Quarterly*, *41*, 126-135. https://doi.org/10.1016/j.ecresq.2017.07.003

Trocmé, N., & Bala, N. (2005). False allegations of abuse and neglect when parents separate. *Child Abuse & Neglect*, *29*(12), 1333-1345. https://doi.org/10.1016/j.chiabu.2004.06.016

Vredeveldt, A., Van Koppen, P. J., & Granhag, P. A. (2014). The inconsistent suspect: A systematic review of different types of consistency in truth tellers and liars. In R. Bull (Eds.), *Investigative interviewing* (pp. 183-207). Springer. https://link-springer-com.qe2a-proxy.mun.ca/book/10.1007%2F978-1-4614-9642-7

Vredeveldt, A., & Wagenaar, W. A. (2013). Within-pair consistency in child witnesses: The diagnostic value of telling the same story. *Applied Cognitive Psychology*, *27*(3), 406-411. https://doi.org/10.1002/acp.2921

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*(4), 239-263. https://doi.org/10.1023/A:1006610329284

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, *26*(3), 261-283. https://doi.org/10.1023/A:1015313120905

Vrij, A., & Mann, S. (2005). Police use of nonverbal behavior as indicators of deception. In R. E.

Riggio & R. S. Feldman (Eds.), *Applications of nonverbal communication* (pp. 63-94).

Lawrence Erlbaum Associates Publishers. https://psycnet.apa.org/record/2005-02117-004

Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006). Detecting lies in young children,

adolescents and adults. *Applied Cognitive Psychology*, *20*(9), 1225-1237.

https://doi.org/10.1002/acp.1278

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies

as a function of police interview styles. Law and Human Behavior, *31*(5), 499-518.

https://doi.org/10.1007/s10979-006-9066-4

Vrij, A. (2008a). Nonverbal dominance versus verbal accuracy in lie detection: A plea to change

police practice. *Criminal Justice and Behavior*, *35*(10), 1323-1336.

https://doi.org/10.1177/0093854808321530

Vrij, A. (2008b). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons Ltd.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal

lie detection. *Psychological Science in the Public Interest*, *11*(3), 89-121.

https://doi.org/10.1177/1529100610390861

Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures:

A comparison between total detail and proportion of complications. *Legal and

Criminological Psychology*, *23*(2), 265-279. https://doi.org/10.1111/lcrp.12126

Warren, K. L., Bakhtiar, A., Mulrooney, B., Raynor, G., Dodd, E., & Peterson, C. (2015).

Adults' detection of deception in children: Effect of coaching and age for children's true

and fabricated reports of injuries. *Behavioral Sciences & the Law*, *33*(6), 784-800.

https://doi.org/10.1002/bsl.2210

Warren, K. L., Peterson, C., & Gillingham, C. C. (2018). Children who are coached to lie: Does

linguistic analysis help in understanding why these children are so

believable? *Psychiatry, Psychology and Law*, *25*(5), 789-805.

https://doi.org/10.1080/13218719.2018.1478336

Wojciechowski, B. W., Gräns, M., & Lidén, M. (2018). A true denial or a false confession?

Assessing veracity of suspects' statements using MASAM and SVA. *PLOS ONE*, *13*(6).

https://doi.org/10.1371/journal.pone.0198211

Zajac, R., & Hayne, H. (2006). The negative effect of cross-examination style questioning on

children's accuracy: Older children are not immune. *Applied Cognitive Psychology*,

20(1), 3-16. https://doi.org/10.1002/acp.1169

Zajac, R., O'Neill, S., & Hayne, H. (2012). Disorder in the courtroom? Child witnesses under

cross-examination. *Developmental Review*, *32*(3), 181-204. doi:

http://dx.doi.org/10.1016/j.dr.2012.06.006

Footnotes

[1] Switching from Survey Monkey to Qualtrics allowed participants to be randomly assigned to conditions, while simultaneously allowing researchers to fill each condition with a certain number of participants.

Appendix A

Assessing the Credibility of Children's Injury Reports

When a child is involved in a criminal case they may have to give testimony to professionals, judges and possibly the court. As in real life, the child may be giving **a truthful account** or **false testimony**. Carefully read the transcript provided and decide whether you believe the child is telling **a fabrication** or **being truthful**. Importantly, note the reasons *why* you came to this conclusion. **Please note** that you may return to the transcript at any point during this study to aid in your decisions.

(see transcript)

Is the child telling the truth (rather than telling a lie)? (Please circle an answer)

Yes   No

How confident are you in your decision? (Please circle an answer)

Not at all   Somewhat   Very confident

1   2   3   4   5

Why did you come to this decision (whether the child was telling the truth, ora lie)? Please indicate any information in the transcript that led you to this decision or any strategies you used in making the decision.

Questionnaire

Please Note: Participants are free to skip any questions they do not wish to answer.

Age: _____

Gender: _____

Ethnicity (Please select all that apply):

Black___

Caucasian ___

Hispanic/ Latino ___

Asian/ Pacific Islander ___

First Nation / Inuit / Metis___

Prefer not to respond____

Not listed (please specify): _____

Current Occupation: _____

Years of experience in current occupation: _____

Please indicate any course work you have completed in forensic psychology, police studies, or criminology. Which of the following have you taken courses in? Please check off all that apply.

___ Forensic Psychology

___ Criminology

___ Training through the RNC or RCMP

___ Social Work

___ Law

____Sheriff's officer

____Probation officer

____Corrections

___ Other (please specify) _____

Please indicate any experience you have with children, including any employment or profession. Please check either yes or no for each experience. Then write the years of experience you have in the right hand column.

Type of Experience:	Years of Experience:
Daycare Worker: ___ Yes ___ No	
Lawyer: ___ Yes ___ No	
Teacher: ___ Yes ___ No	
Social Worker: ___ Yes ___ No	

Police Officer: ___ Yes ___ No	
Babysitter: ___ Yes ___ No	
Coach: ___ Yes ___ No	
Team Leader: ___ Yes ___ No (ex: Scouts)	
Parent or Guardian: ___ Yes ___ No	
Other: ___ Yes ___ No Please Specify:_____	

Highest level of education completed (please check one):

___No education

___ Junior high school

___High school diploma or GED

___Post-secondary certificate/diploma

___ University 1 year to 3 years

___ Bachelor's degree

___Graduate school or other advanced degree

Are you currently a student at a University or College? (Please check one)

___Yes

___No

What is your household average yearly income? (Please check one)

___$\leq$ $20,000

___ $20,001 – $40,000

___ $40,001 - $60,000

___ $60,001- $80,000

___ $80,001 - $100,000

___ $\geq$ $100,001

Is English your native language?

___ Yes   ___ No

At least what percentage of time would you say you speak English in your home? (Please check one)

___Less than 50% of the time

___ Between 51% - 75% of the time

___ Greater than 75% of the time

Have you ever experienced an injury or been through an experience similar to the child in the interview you just read?

___ Yes   ___ No

If yes, explain: